

Интегрированное управление качеством данных в организации

О. И. Захарова, В. С. Коробейников

Поволжский государственный университет телекоммуникаций и информатики (ПГУТИ)

Аннотация: Актуальность задачи интегрированного управления качеством данных возрастает в условиях увеличения объемов, разнообразия и критичности используемых данных. Несмотря на это, в организациях есть пробелы в понимании взаимосвязей между качеством данных, качеством процессов и информационных систем. Целью настоящего исследования является систематический анализ существующих методологий и концепций управления качеством данных, а также выявление ключевых проблем при их внедрении. В работе проведен обзор научных источников, представлены стандартные элементы и схема интегрированного подхода к качеству данных. На основе архитектуры Data Lakehouse разработана высокоуровневая схема потоков данных, отражающая взаимодействие компонентов системы. Обоснована необходимость разработки новых методов и алгоритмов для оптимизации качества больших данных, выходящих за рамки традиционных парадигм, ориентированных на структурированные данные. Определены и систематизированы ключевые проблемы, часто игнорируемые на практике, и сформированы критерии для успешного внедрения интегрированного подхода к управлению качеством данных.

Ключевые слова: качество данных, управление качеством данных, интегрированное управление качеством данных, методология качества данных, мониторинг качества данных, большие данные, качество информации.

Для цитирования: Захарова О. И., Коробейников В. С. Интегрированное управление качеством данных в организации // Вестник СибГУТИ. 2025. Т. 19, № 4. С. 28–47. <https://doi.org/10.55648/1998-6920-2025-19-4-28-47>.



Контент доступен под лицензией
Creative Commons Attribution 4.0
License

© Захарова О. И., Коробейников В. С., 2025

Статья поступила в редакцию 24.09.2025;
переработанный вариант – 30.09.2025;
принята к публикации 21.10.2025.

1. Введение

В современном мире ввиду стремительного роста объема информации ее качество является важнейшим аспектом для принятия решений. Согласно исследованию Ихуа Шэн о взаимосвязи качества информации с ценностью услуг и продукции, высокое качество данных положительно связано с эффективностью компании [1].

Определение качества данных имеет множество значений, которые связаны с контекстом данных, доменом и областью, в которых они используются. В рамках данной работы под качеством данных понимается совокупность характеристик, отражающих их пригодность для достижения конкретной цели [31]. Качественные данные обеспечивают уверенность в том, что принимаемые на их основе решения являются достоверными и надежными. Напротив, ошибки и неточности в данных или связанных с ними процессах могут привести к искаженным выводам, ошибочным управленческим решениям и снижению операционной эффективности [34].

В связи с этим системный подход к обеспечению качества данных на всех этапах их жизненного цикла, объединяющий методы оценки, технологии и процессы, приобретает критически важное значение для контроля и поддержания приемлемого уровня качества вне зависимости от цели и сферы применения данных. Такой подход называется интегрированным управлением качеством данных [26] и позволяет повысить ценность данных, эффективность принятия решений, а также снизить риски и издержки.

Таким образом, целью настоящей статьи является исследование концепций и подходов к интегрированному управлению качеством данных, а также выявление проблем, возникающих при его внедрении и применении для оценки качества данных и связанных с ними процессов.

2. Обзор литературы

Проблематика интегрированного управления качеством данных начала активно рассматриваться в научном сообществе начиная с конца прошлого века. Так, в статье Ричарда Уанга был предложен первый общий методологический подход к полному управлению качеством данных – TDQM (Total Data Quality Management) [2]. В своей работе автор представил системный подход к управлению качеством в рамках информационных систем, включающий определение, измерение, анализ и улучшение качества информации. Согласно методологии TDQM, для эффективного управления качеством необходима вовлеченность всех заинтересованных сторон, включая разработчиков и бизнес-пользователей. При этом организации должны рассматривать данные как продукт системы производства информации подобно физическому продукту, учитывая особенности информационных продуктов. В TDQM применяется язык моделирования IP-MAP (Information Production Map), позволяющий описывать жизненный цикл данных и управлять их качеством на каждом этапе. Впоследствии IP-MAP был расширен до IP-UML для управления растущей сложностью информационных систем. Основные этапы методологии адаптированы под цикл Шухарта-Деминга (PDCA – Plan-Do-Check-Act), обеспечивая итеративный процесс управления качеством данных:

1. Планирование. С использованием IP-MAP определяются требования к качеству информации с позиции различных участников – поставщиков, производителей, потребителей и владельцев информационных процессов.
2. Измерение. Проводится оценка качества данных по ключевым характеристикам (полнота, точность и пр.).
3. Анализ. Выявляются причины несоответствий и проблем в данных.
4. Улучшение. На основе полученных метрик разрабатываются и внедряются меры по улучшению качества данных.

Методология TDQM оказала значительное влияние на формирование последующих концепций, международных стандартов и практик в области управления с данными. Она значительно повлияла на развитие Data Governance – корпоративного управления данными.

Исследователи Ян В. Ли, Д. Стронг, Беверли К. Кан, Ричард Уанг предложили концепцию оценки качества информации AIMQ (Methodology for Information Quality Assessment) [3], которая согласуется с основными принципами TDQM и дополняет их. AIMQ рассматривается как управленческий инструмент для концептуализации и оценки качества информации с позиции бизнес-пользователей, разработчиков и менеджеров. Концепция применяет бенчмарк-анализ и может использоваться в различных предметных областях, позволяя выявлять приоритетные направления улучшения качества с использованием модели PSP/IQ (Product-Service-Performance/Information Quality). Оценка качества в AIMQ структурирует оценку с помощью анкетирования по следующей структуре:

1. Пилотный опрос. Используется для выявления значимых характеристик качества, которые следует сравнивать.
2. Основной опрос. Направлен на получение количественных оценок по выявленным характеристикам и атрибутам качества.

3. Сравнение результатов с эталоном (бенчмарком). Проводится анализ разрывов (Gap Analysis) для интерпретации различий между организациями, демонстрирующими лучшие практики, и ожиданиями пользователей внутри компании. По результатам анализа формируется представление о состоянии управления качеством данных организации и разрабатываются рекомендации по его улучшению.

Исследователи Батини К. и Сканныпьеко М. разработали методологию для оценки и улучшения качества данных CDQ (Complete Data Quality), которая помогает выбрать оптимальный процесс улучшения качества данных с учетом исходных требований и позволяет количественно оценить в какой степени бизнес-процессы подвержены применению некачественных данных [5]. Основные этапы методологии:

1. Восстановление информации. На этом этапе устанавливаются взаимосвязи между процессами, сервисами и данными. Эти связи моделируются с помощью матриц, описывающих подразделения пользователей данных и их роль в бизнес-процессах, нормативные и организационные правила, процессы и их зависимость от качества данных.

2. Оценка. Проводится анализ качества информационных потоков и данных по характеристикам качества. Дополнительно рассчитывается оценка затрат и выгод по сформированным новым целевым уровням качества, необходимым для повышения эффективности процессов.

3. Улучшение. Определяются причины ошибок и способы их решения. Выбор оптимального процесса улучшения и целевых уровней качества производятся по оценке наилучшего соотношения эффективности, затрат и выгод.

Кроме описанных методологий, на этапе развития области управления качеством данных были предложены и иные концепции [6]:

– Авторы Кен-Сок Рю, Джу-Сок Пак и Джэ-Хон Пак предложили модель зрелости управления качеством структурированных данных DQMMM (Data Quality Management Maturity Model) [4]. Данная концепция описывает подход на основе четырех ключевых направлений: интеграции баз данных, архитектуры данных, структуры метаданных и стадии зрелости управления данными. DQMMM предполагает наличие уровней зрелости от начального, при котором качество данных не контролируется, до оптимизированного, при котором реализован системный и устойчивый процесс управления качеством. По результатам оценки текущего состояния управления качеством в указанных направлениях разрабатывается план по улучшению каждой из областей с целью повышения общей эффективности компании и процесса управления данными.

– Йесфельд М., Квикс К. и Ярке М. разработали концепцию DWQ (Data Warehouse Quality), направленную на обеспечение качества при проектировании и эксплуатации хранилищ данных [7]. Методология охватывает три направления: концептуальную бизнес-перспективу с фокусом на бизнес-модели организации; логическую перспективу, отражающую структуру хранилища данных; физическую перспективу, охватывающую физический слой передачи данных. Для каждого направления формируется оценка данных по характеристикам качества. С точки зрения этапов управления качеством структура DWQ схожа с методологией TDQM, однако стадия улучшения не включает подробного описания методик и рекомендаций.

– Чапман А., Ричардс Х. и Хоукен С. представили CIHI (Canadian Institute for Health Information), ориентированную на оценку качества больших и неоднородных баз данных в сфере здравоохранения [8]. Методология включает в себя две фазы:

1. Разработка рамочной модели качества данных, содержащей три этапа: стандартизацию информации о качестве данных; разработку стратегии оценки; определение процесса управления данными, включая разработку программы непрерывного улучшения и приоритизации задач.

2. Проведение глубокого анализа наиболее часто используемых данных. Включает в себя три этапа: анализ; оценку; документирование результатов. Оценка осуществляется с использованием иерархической модели, где первый уровень – это анкеты на основе 86 базовых

критериев; второй уровень – агрегирование результатов первого уровня в 24 характеристиках качества с использованием опеределенных алгоритмов; третий уровень – агрегация характеристик в пять обобщенных категорий – точность, своевременность, сравнимость, пригодность для использования и актуальность; четвертый уровень – формирование единой итоговой оценки качества. На каждом уровне проводится валидация для обеспечения единообразия и сопоставимости интерпретации результатов оценки.

– М. Дж. Эпплер, П. Мюнценмайер разработали IQM (Information Quality Measurement). IQM, одну из первых концепций, ориентированных на обеспечение качества веб-данных [9]. Концепция предлагает два аспекта: рамочную модель качества, включающую характеристики качества; план действия, описывающий процесс проведения оценки. Основные этапы концепции:

1. Рамочная структура формирует список характеристик качества, полученных на основе требований пользователей.
2. Проводится валидация данных, в ходе которой каждая характеристика сопоставляется с методом оценки.
3. Оценка качества по выбранным критериям.
4. Анализ результатов и определение корректирующих действий, включающих контроль качества и исправление ошибок.

– Фалорси П., Паллара С., Павоне А., Алессандрони А., Масселла Э. и Сканнапьеко М. представили ISTAT (Italian National Bureau of Census), разработанную в рамках национального бюро переписи населения Италии [10]. Методология концентрирует внимание на учете гетерогенности источников, анализируя в дополнение к упоминаемым ранее характеристикам качества сопоставимость данных, которую можно рассматривать как частный случай согласованности. Основные этапы концепции:

1. Оценка. Проводится анализ данных ISTAT с целью выявления проблем с качеством по характеристикам, с использованием статистических методов;
2. Глобальное улучшение. Выполняется связывание записей между региональными базами данных и разрабатываются методы по улучшению данных. Интеграция данных происходит через событийно-ориентированную архитектуру (event-driven);
3. Региональное улучшение. Выполняется улучшение качества на уровне отдельных региональных источников;
4. Совместное улучшение. Проводится координация различных источников данных с целью оптимизацию процессов обмена и приведения данных к единому формату.

– Су Ю. и Джин З. представили AMEQ (Activity-based Measuring and Evaluating of Product information Quality), основной целью которой является обеспечение надежной основы для оценки и улучшения качества данных о продукте в соответствии с целями организации [11]. Данная концепция ориентирована на компании, в которых информация о продуктах – это ключевой компонент баз данных. Основные этапы концепции:

1. Оценка организационной зрелости бизнес-процессов и определение характеристик качества данных, классифицируемых по значимости бизнес-процессов;
2. Спецификация информационного продукта. Сопоставление бизнес-процессов и данных с применением объектно-ориентированного подхода ООА (Object-oriented approach). Разрабатывается модель измерения качества по 8 типам объектов: человеческие, информационные, входные, выходные и процессные ресурсы, виды деятельности, цели и метрики эффективности предприятия компании.
3. Оценка качества данных проводится на основе собранных критериев;
4. Анализ причин проблем с качеством. Выявление первопричин проблем с качеством по общим рекомендациям AMEQ;
5. Улучшение. Способы улучшения определяются в зависимости от степени влияния на результативность бизнес-процесса. Методология предлагает только общие рекомендации без конкретных инструментов и методов.

– Лошин Д. разработал COLDQ (Cost-Effect Of Low Data Quality), которая описывает систему количественной оценки влияния низкого качества данных на бизнес-процессы [12]. Основные этапы концепции:

1. Построение карты информационных потоков. Включает в себя два типа моделей: стратегический для принятия решений и операционный для обработки данных;
2. Анализ качества. Заключается в применении количественных и качественных методов, включая интервью с внутренними и внешними пользователями, с целью обнаружения ошибок в данных и процессах;
3. Идентификация ошибок. Обнаруженные ошибки и неточности в данных связываются с конкретными действиями стратегической и операционной моделей;
4. Оценка затрат. Классификация экономических последствий от использования некачественных данных проводится по следующим категориям: операционные, тактические и стратегические.
5. Улучшение. Анализ затрат применяется для выбора мероприятий по улучшения качества данных.
6. Анализ рентабельности. Заключается в оценке точки безубыточности (break-even points) и возврата инвестиций (ROI), что позволяет обосновать принимаемые меры с точки зрения конкретных финансовых показателей.

– Сканнапьеко М., Виргилито А., Маркетти М., Мечелла М. и Бальдони Р. разработали DaQuinCIS (Data Quality in Cooperative Information Systems), которая рассматривает вопросы качества в условиях межорганизационного взаимодействия [13]. Авторы уделяют особое внимание использованию метаданных о происхождении источников, а также механизмам выбора наиболее качественных данных при наличии дублирующихся источников. DaQuinCIS предлагает использовать модель D2Q (Data and Data Quality), которая включает: сами данные; свойства качества данных; степень доверия к источнику, оцениваемая третьей стороной; связь между данными и метаданными о их качестве. С точки зрения этапов управления качеством структура схожа с методологией TDQM, однако на этапе улучшения данные либо снова проходят цикл оценки и улучшения, либо считаются пригодными для дальнейшей передачи. DaQuinCIS включает следующие особенности:

1. Архитектура методологии ввиду спецификации делится на: внутреннюю, включающую фабрику качества данных для каждой организации; внешнюю, обеспечивающую межорганизационную коммуникацию и содержит сервис уведомлений о текущем состоянии качества;
2. Функционирование архитектуры организовано в зависимости от типа запросов: пользовательские направлены на получение данных и формирование требований к их качеству; системные применяются для оценки качества с последующим уведомлением о результатах оценки; направленные на улучшение и обрабатываемые модулем оценки качества с применением алгоритма Record Matcher, позволяющего сопоставлять и отбирать лучшие версии записей на основе заданных характеристик качества.

– Амицис Ф. и Батини К. представили методологию QAFD (Quality Assessment of Financial Data), направленную на определение стандартных мер качества операционных финансовых данных с целью минимизации затрат на инструменты оценки [14]. Основные этапы концепции:

1. Выбор наиболее значимых финансовых атрибутов. Отбор проходит на основании предыдущих оценок и критичности данных, а переменные группируются в отдельные категории в зависимости от их типа данных, влияния на потребителей, бизнес-процессов и пр.;
2. Анализ. Формируются правила проверок по характеристикам качества: точность, полнота, согласованность, уникальность, своевременность;
3. Количественная оценка. Авторы предлагают алгоритмический подход, позволяющий рассчитать оценку и провести ранжирование по каждой характеристике качества;

4. Качественная оценка. Проводится с позиции бизнес-экспертов, клиентов и специалистов по качеству данных. Итоговая оценка формируется как среднее значение по каждой группе экспертов;

5. Сравнение качественной и количественной оценки. Если разница между оценками положительная, то количественная оценка подлежит пересмотру с целью выявления проблем с качеством по мнению экспертов.

Дальнейшее развитие области интегрированного управления качеством данных основывается на вышеприведенных методологиях. Среди существующих современных концепций отмечены следующие:

– Концепции зрелости работы с большими данными BDMM (Big Data Maturity Model) [27], применяемые для оценки зрелости организации в части работы с данными. BDMM концепции предполагают поэтапное внедрение компонентов управления данными, охватывающих как структуру, так и процессы. BDMM следует рассматривать как дополнительный инструмент для оценки и корректировки стратегии управления данными в организации. BDMM концепции предполагают наличие 5 уровней зрелости:

1. Начальный, где процессы не контролируются.
2. Управляемый, где процессы существуют, но являются реактивными. На этом этапе начинается стандартизация процессов, данные используются для определенных целей.
3. Определенный, где процессы стандартизированы и обычно документируются. На данном этапе используются инструменты для детального анализа данных, инфраструктура масштабируется, а также происходит внедрение моделей машинного обучения.
4. Количественно управляемый, где процессы измеряются и контролируются. На текущем этапе использование данных интегрировано в бизнес-процессы, а эффективность всех процессов предсказуема и управляема.

5. Оптимизированный, где процессы направлены на постоянное совершенствование.

– Концепция IBM DGF (Data Governance Framework) [29], которая основывается на сводке лучших практик управления качеством данных IBM DGI (Data Governance Institute Framework) [28] и является ее адаптацией под корпоративное управление. Данный подход включает в себя определение проблем, оценку зрелости и разработку дорожной карты по развитию управления качеством. Кроме этого, предлагает способы и методы оценки качества и характеристики данных, упоминаемые в ранних методологиях.

– Методология DQVM (Data Quality Validation Methodology) [35] авторов Каппиелло Ч. и др., разработанная для оценки влияния качества данных на бизнес-процессы. DQVM предлагает методику валидации качества данных, которая позволяет экспертам назначать веса характеристикам качества в зависимости от их значимости для конкретного процесса. Концепция включает анализ отклонения показателей между корректными и ошибочными данными для оценки их влияния на бизнес-процесс.

– Концепция BDQMF (Big Data Quality Management Framework) [33] авторов Талей И. и др., разработанная для управления качеством систем больших данных. Данная концепция сконцентрирована на распределенных и масштабируемых архитектурах систем хранения и обработки больших данных. Основные этапы BDQMF:

1. Проект качества больших данных. Включает в себя различные источники, модель данных, требования к качеству от бизнес-пользователей в зависимости от бизнес-процессов.
2. Оценка. Заключается в использовании DQP (Data Quality Profile), который включает в себя профиль данных и качества, а также репозиторий для последующего отслеживания и управления. Оценка состоит из следующих этапов:
 - а) подготовка и первоначальное профилирование данных;
 - б) профилирование качества с целью выявления закономерностей в данных;
 - в) определение ключевых характеристик качества данных;
 - г) количественная оценка;
 - д) проверка на обеспечение соответствия данных установленным стандартам;

- f) определение правил оценки качества данных, которые могут быть использованы для корректировки и улучшения;
- g) проверка соответствия правил стандартам и самим данным;
- h) улучшение и корректировка правил.

3. Обработка данных. Состоит из предобработки и обработки данных в соответствии с установленными стандартами. Каждое сформированное правило оценки представляет из себя функции предварительной обработки, однако перед их применением DQP настраивается и пересматривается экспертами в соответствии с эталонным отчетом оценки качества данных.

4. Контроль качества. Проводится мониторинг на всех этапах жизненного цикла для выявления и устранения проблем с качеством.

Кроме приведенных концепций вклад в развитие области внесло руководство по управлению данными DAMA-DMBOOK [15], которое включает в себя набор рекомендаций и лучших практик по аспектам управления данными. Также существенный вклад в развитие интегрированного управления качеством данных внесли международные стандарты серии ISO/IEC, среди которых:

- ISO 38500 и ISO/IEC 38505-1:2017, предоставляющие структуру для обеспечения эффективного управления данными, включая цели, роли, оценку рисков и формирования отчетности по данным и их ключевым характеристикам. Российский стандарт ГОСТ Р ИСО/МЭК 38500-2017 идентичен ISO/IEC 38500:2015;

- ISO 8000, описывающий процесс управления и требования к качеству критически важных данных [16]. Российский стандарт ГОСТ Р ИСО 8000-8-2022 идентичен ISO 8000-100:2016;

- ISO/IEC 11179, который описывает метаданные и подходы к их управлению с целью обеспечения единообразия и совместимости данных [17]. Российский стандарт ГОСТ Р ИСО/МЭК 11179-1-2010 идентичен ISO/IEC 11179-1:2004;

- ISO/IEC 20943, описывающий обмен информацией между системами и их семантическую согласованность, а также аспекты больших данных [18];

- ISO/IEC 25012:2008, устанавливающий модель качества продуктов программного обеспечения и качества данных с точки зрения их характеристик [19]. Данный стандарт включает в себя 15 характеристик качества данных и делит их на внутренние и внешние;

- ISO/IEC WD 5259 [20, 21], разрабатываемая в настоящее время серия стандартов для управления качеством в рамках задач аналитики и машинного обучения. Стандарт описывает жизненный цикл управления качеством данных DQMLC (data quality management life cycle), включает рекомендации и практики в области стратегического управления качеством из более ранних методологий и стандартов серии ISO/IEC.

Анализ источников показывает, что интегрированное управление качеством данных является важнейшей составляющей всеобъемлющего процесса Data Governance и охватывает различные технические, экономические, юридические и гуманитарные аспекты. К техническим аспектам относится разработка архитектуры и инфраструктуры для автоматизации, оценки качества данных и процессов. Экономические включают в себя оценку влияния низкого качества на процессы, анализ затрат и выгод от улучшения качества, расчет окупаемости инвестиций в управление. К гуманитарным и юридическим относятся организационные, управленческие и этические аспекты формирования культуры работы с данными, распределения обязанностей и ролей, соблюдения нормативно-правовых требований.

3. Материалы и методы

Для достижения исследовательской цели авторами используются методы анализа, синтеза и систематизации информации касательно концепций к интегрированному управлению качеством данных. Рассмотрим сформированную на основании описанных концепций классификацию (табл. 1).

Таблица 1. Классификация концепций управления качеством данных

№	Категория	Концепция\Стандарт
1	Комплексный охват аспектов	CDQ
2	Управленческие\стратегические структуры	IBM DGF, DAMA-DMBOOK, ISO 8000, ISO 38500, ISO/IEC WD 5259, BDMM
3	Оценка качества, метрики. Концентрация на оценке качества, определении критериев и показателей; частично охватывают улучшения.	AIMQ, CIHI, AMEQ, QAFD, IQM, DQVM, ISO/IEC 25012
4	Операционный подход. Направлены на техническую реализацию контроля и обеспечения качества, управление метаданными и процессы разнообразности данных.	TDQM, DWQ, ISTAT, DaQuinCIS, DQMMM, BDQMF, ISO/IEC 11179, ISO/IEC 20943
5	Экономические аспекты. Фокус на оценке затрат на внедрение и поддержание управления качеством данных.	COLDQ

С учетом приведенной выше классификации отмечено, что экономические аспекты рассматриваются менее активно. Это может быть обусловлено тем, что преимущества внедрения управления качеством данных очевидны, а оценка затрат и выгод проводится при наличии актуальных показателей качества. В то же время практики по оценке эффективности внедрения и поддержания систем управления данными представляют собой обширную область, выходящую за рамки данного исследования.

Юридические и гуманитарные аспекты, напротив, включают в себя современные стандарты и практики управления данными, такие как DAMA-DMBOOK. Юридические вопросы вызваны изменяющейся нормативно-правовой базой и активным развитием области машинного обучения, что требует соблюдения этических норм, стандартов и требований регуляторов. Гуманитарные аспекты охватывают структуру организации и управления, формирование культуры качества данных, включающую распределение ролей, обязанностей и развитие компетенций сотрудников. Эти два аспекта часто представляют собой проблемные места при внедрении интегрированного управления качеством данных, поскольку требуют значительных усилий при трансформации управленческой структуры и поведенческих установок сотрудников.

Технические аспекты оценки качества данных и процессов, включая методологию и методы, претерпевают изменения, поскольку большинство концепций, сформировавшихся на первоначальном этапе развития области управления качеством данных, основаны на традиционных стратегиях обработки структурированных данных. Однако в современном мире, в связи с ростом объема информации и разнообразия данных, архитектурные решения, инструменты оценки, мониторинга и улучшения качества данных должны быть адаптированы для работы с большими данными. Разрабатываемые международные стандарты серии ISO/IEC WD 5259 придерживаются этой позиции. Среди современных концепций выделяется BDQMF, которая является современным подходом к оценке качества больших данных и включает в себя требования и цели применения данных, целевые показатели характеристик качества, а также модели данных и разнообразные источники: устройства интернета вещей (IoT), хранилища данных (DWH), базы данных, веб-данные, логи, видео, аудио, облачные сервисы и другие.

Большинство авторов рассмотренных концепций сходятся во мнении, что качество данных связано с этапами их жизненного цикла. В связи с этим управление качеством данных должно быть встроено в каждый из этих этапов и основываться на реализации классического цикла Шухарта-Деминга. На рисунке 1 представлен обобщенный принцип интегрированного управления качеством данных.



Рис. 1. Обобщенный цикл интегрированного управления качеством данных

Источник: составлено автором

Интегрированное управление качеством данных начинается с этапа планирования (Plan), в рамках которого определяются и приоритизируются задачи в соответствии с оценкой текущего уровня качества данных и процессов. На этапе реализации (Do) проводятся мероприятия по устранению первопричин проблем, формируются требования к последующей оценке и мониторингу. На стадии проверки (Check) осуществляется мониторинг качества данных по установленным бизнес-требованиям и допустимым пределам качества. На этапе корректировки (Act) проводятся оперативные работы по устранению текущих проблем с данными с целью повышения качества. Перезапуск цикла управления качеством осуществляется при увеличении объема и критичности инцидентов; выходе установленных диапазонов качества за допустимые пределы; появлении новых наборов данных; изменении бизнес-правил, требований или стандартов.

Область применения и тип информации влияют на планирование проверок и мероприятий по обеспечению качества данных. Выделяются следующие подходы:

1. Контентно-ориентированные (Content-based), определяющие качество по самим данным. Например, AIMQ, CDQ, DQVM, BDQMF (также учитывает работу с метаданными);
2. Контекстно-ориентированные (Context-based), которые определяют качество по метаданным, описывающим структуру, содержание, качество и местоположение информации. Например, TDQM, DaQuinCIS, DQMMM, BDMM.

Вне зависимости от используемого подхода, оценка качества данных, как правило, базируется на анализе их характеристик. Набор характеристик в описанных концепциях вариативен, что подтверждает многомерность структуры оценки качества. Однако все исследователи сходятся во мнении, выделяя важные базовые характеристики: точность, полнота, согласованность и своевременность. Несмотря на то что характеристики качества – это важная составляющая интегрированного управления качеством данных, детальное описание каждой из них выходит за рамки настоящего исследования и будет являться целью последующих работ.

Выделяются следующие подходы к оценке качества:

1. Количественная оценка (Quantitative assessment). Данный подход основан на измерении числовых показателей в соответствии с экономической целесообразностью или характеристиками качества данных – полнота, точность, согласованность, своевременность и пр.

2. Качественная оценка (Qualitative assessment). Данный подход основан на субъективных характеристиках качества данных, таких как понятность, полезность и других.

3. Комбинированная оценка (Mixed assessment). Включает в себя количественную и качественную оценку.

Процесс получения оценок качества включает в себя различные методы и подходы к оценке. В рамках данного исследования все они выделены в следующие группы:

1. Профилирование данных (Data profiling). Профилирование данных – это процесс изучения данных с целью глубокого понимания самих данных и их качества. Этот подход включает создание статистических оценок в разрезе характеристик качества и/или метаданных, представляющих детальную физическую, логическую и концептуальную информацию. Выделяют следующие виды:

- a. Онлайн-профилирование, осуществляемое в реальном времени при поступлении новых данных;
- b. Инкрементальное профилирование, добавляющее новую информацию к существующему профилю данных по мере их поступления;
- c. Структурное профилирование, собирающее информацию по структурам и схемам данных;
- d. Непрерывное профилирование, направленное на постоянный процесс мониторинга и обновления профиля данных.

Профилирование данных в общем случае заключается в поиске закономерностей (примитивов) в данных. Поиск закономерностей достаточно трудоемкая в вычислительном плане задача. Существует различные примитивы, вот некоторые из них:

- a. Функциональные зависимости. Включают в себя большой класс всевозможных задач, например, поиск связанных колонок и опечаток в тексте;
- b. Ассоциативные правила. Применяются к транзакционным данным и логам, включают правила вида «если в данных есть А, то с некоторой вероятностью ожидается и В». Также может быть применимо для выбора правил качества;
- c. Условные функциональные зависимости. Является обобщением приведенных примитивов, выполняемых на подмножестве данных.

2. Валидация данных и процессов (Data and process validation). Заключается в подтверждении того, что данные\метаданные\процессы надежны и соответствуют определенным бизнес-правилам или заданным спецификациям и стандартам. Данный подход также можно сопоставлять с характеристиками качества.

Понятие валидации и профилирования данных частично пересекаются, а их границы не всегда возможно строго определить. Методы мониторинга, использующие эти способы оценки, по степени реализации сгруппированы следующим образом:

1. Алгоритмически-ориентированные (Algorithm-based). Применяет различные формализованные концептуальные и алгоритмические структуры для оценки. Концепции DQVM, DQMMM, BDMM и др. используют данный подход;

2. Основанные на правилах (Rule-based). Правила включают установление ограничений и необходимы для решений и/или предотвращения проблем с качеством. Как правило, используется в рамках концепций, ориентированных на данные, и включают в себя:

- a. Бизнес-правила (Business DQ Rules), которые описывают требования к качеству в терминах понятных бизнесу с указанием на процессы, где они применяются;
- b. Техническая спецификация правил (DQ Rule Specification), уточняющая как реализуется проверка каждого правила.

3. Модельно-ориентированные (Model-based). Включает статистические и экономические модели, а также алгоритмы машинного обучения для выявления аномалий, паттернов и правил в данных. Частично применяется в концепции BDQMF для определения возможных правил с помощью алгоритма Random Forest и XGBOOST.

При наличии неудовлетворительных результатов оценки качества данных предполагаются мероприятия по улучшению данных, которые заключаются в двух подходах:

1. Ориентированные на данные (Data-driven). Улучшение качества данных производится на этапе обработки данных, а также исправляются исторические ошибки в данных. Например, AIMQ, CDQ, DQVM, BDQMF;

2. Ориентированные на процессы (Process-driven). Улучшение качества данных применяется на каждом этапе создания данных, а также исправляются проблемы в бизнес-процессах, приведшие к ошибкам. Например, TDQM, CDQ, BDQMF, DaQuinCIS, DQMMM, BDMM.

В части мониторинга качества данных необходимо отметить актуальную на момент написания статьи классификацию по степени автоматизации проверок. Данная классификация не противоречит описанным выше подходам, напротив, включает в себя решения, которые их комбинируют.

1. Data Testing [22]. Концепция отражает тестирование данных по определенным правилам на предмет проблем до их распространения в другие наборы данных в качестве источников. Концепция определяется низким покрытием и неэффективностью при большом количестве различных источников данных ввиду невозможности определения первопричин ошибок и ограничений в масштабируемости. Data Testing подходит для компаний начального уровня зрелости. Включает в себя базовые проверки на качество данных в разрезе характеристик, методы валидации и профилирования данных.

2. Data Quality Monitoring [23]. Процесс автоматического отслеживания качества данных, который для выявления аномалий использует модельно-ориентированные подходы. Концепция обеспечивает большее покрытие, чем Data Testing. Алгоритмы машинного обучения в этом подходе способны устанавливать пороговые значения, нет необходимости трудного переноса проверок с одного набора данных на другой. К основным ограничениям подхода относятся сложность работы с неизвестными (новыми) наборами данных и затруднения в выявлении корневых причин проблем с качеством. Data Quality Monitoring подходит для компаний определенного уровня зрелости.

3. Data Observability [24]. Концепция позволяет организации лучше понимать состояние данных путем выявления аномалий в процессах, системах, инфраструктуре, а также обеспечивает быстрое реагирование на возникающие инциденты. Подход компенсирует ограничения Data Testing и Data Quality Monitoring, поскольку обеспечивает мониторинг и поиск сбоев потоков данных и бизнес-процессов на более широком уровне чем использование метрик, логов и трассировок. Основное отличие концепции от ее предшественников не в технологиях, а в самом подходе и инструментах. Смысл этой концепции состоит в выявлении времени простоя данных (Data Downtime), которое означает что произошел инцидент в процессах, приведший к изменению данных. Data Observability требует большей зрелости инфраструктуры и является сложным к внедрению инструментом.

Data Observability предлагает продвинутые инструменты с использованием подхода к автоматическому выявлению правил. Однако, согласно работе Никифоровой А. и Хейди К. из 151 проанализированного инструмента на 2024 год лишь 10 из них заявляют о возможности автоматического выявления правил [30]. Эти инструменты используют различные подходы, которые генерируют правила на основе Model-based, включая использование обучения без учителя и гибридные подходы с ручной настройкой. Однако, полностью автоматической генерации бизнес-правил без участия пользователя не поддерживает ни один инструмент, а также не все они применимы для использования в хранилищах больших данных.

4. Результаты и обсуждение

В настоящее время не существует универсального подхода к интегрированному управлению качеством данных. Это связано с многообразием сфер применения данных, каждая из которых предъявляет различные требования ко всем описанным аспектам управления качеством. В рамках настоящего исследования проанализировано, какие элементы интегрированного управления качеством данных должны учитываться независимо от области применения данных (табл. 2).

Таблица 2. Стандартные элементы интегрированного управления качеством данных

№	Направление	Элементы
1	Стратегия	<p>Политика и стратегия качества данных, включая единое определение характеристик качества данных и их мер (Data Strategy).</p> <p>Управление в соответствии с бизнес-процессами и целями компании.</p> <p>Ключевые показатели эффективности качества (KPI).</p> <p>План по последовательному непрерывному улучшению качества данных и процессов, учитывая оценку уровня зрелости компании.</p> <p>Оценка рисков, затрат и выгод от инициатив по улучшению качества.</p> <p>Анализ влияния качества на результативность компании.</p> <p>Контроль и обеспечение качества данных с учетом их жизненного цикла.</p> <p>Использование Content-based и Context-based подходов к оценке качества.</p> <p>Process-driven\Data-driven подход улучшения качества данных.</p>
2	Организация	<p>Методология управления качеством, включая соглашение об обеспечении качества SLA (Service Level Agreement).</p> <p>Описание повседневных практик и процессов управления данными (Data Operations).</p> <p>Роли и их взаимодействие (Chief Data Officer, Data Steward, Data Owner, Data Custodian).</p> <p>Определение информационных продуктов и их источников, форматов и владельцев.</p> <p>Обучение сотрудников.</p> <p>Формирование культуры организации, ориентированной на данные.</p> <p>Выстраивание взаимодействия в части данных с другими организациями.</p> <p>Проведение регулярных аудитов для выявления проблем и поддержания актуальности стратегии качества данных.</p> <p>Обеспечение защиты персональных данных, доступа и соответствия регуляторным требованиям.</p>
3	Технологии	<p>Определение целевой системы хранения данных с возможностью масштабирования (DWH, Data Lake/Data Lakehouse/Data Fabric/Data Mesh [25]).</p> <p>Методы очистки, профилирования, валидации и мониторинга данных и процессов с включением практик DataOps (Data Quality Monitoring и Data Observability).</p> <p>Процесс обнаружения и исправления ошибок в данных (Data Cleansing).</p> <p>Процесс управления метаданными, документирование информации о происхождении и изменении данных (Data Provenance и Data Linage).</p> <p>Разработка и развитие инструментов для оценки и мониторинга качества данных.</p> <p>Разработка и поддержание централизованного хранилища метаданных организации (Data Catalog).</p> <p>Процесс оценки программного обеспечения на соответствие требованиям (Data assurance).</p>

Внедрение интегрированного управления качеством данных долгосрочный процесс, который начинается с формирования стратегии. В соответствии с ней разрабатывается архитектура и методология управления качеством данных в организации. Далее осуществляется

итеративный, повторяющийся процесс на основе цикла Шухарта-Деминга, при этом управление качеством данных осуществляется сверху вниз и включает в следующие этапы:

1. Сбор бизнес-требований от потребителей данных (Data Users).
2. Формирование правил качества на основе полученных требований.
3. Конкретизация и утверждение правил ответственными лицами (Data Stewards).
4. Внедрение проверок в хранилище данных.

Исходя из вышеописанного, схема интегрированного управления качеством в общем виде представляет из себя базовый сценарий, применимый вне зависимости от сферы использования данных (рис. 2).



Рис. 2. Схема интегрированного управления качеством
Источник: составлено автором

Приведенная схема интегрированного управления качеством данных включает итеративный процесс. В рамках данной схемы предусмотрена возможность корректировки архитектуры качества данных, что может требоваться в условиях изменения регуляторных требований или перехода на новые системы хранения данных. В то же время на практике более эффективно и экономически целесообразно закладывать возможность масштабирования архитектуры на этапе ее проектирования.

В части оценки качества данных целесообразно объединять контентно-ориентированные (Content-based) и контекстно-ориентированные (Context-based) проверки, мониторинг и улучшение качества данных в единую адаптивную платформу, которая позволяет динамически выбирать подходы в зависимости от значимости данных, бизнес-процессов и текущих

оценок характеристик качества, а также формировать и адаптировать правила качества. Описанный подход к оценке приведен на рис. 3.



Рис. 3. Цикл оценки качества данных

Для поддержания высокого уровня точности, согласованности и надежности данных в организации необходимо системно оценивать данные, планировать улучшения и развивать технологии для автоматизации [32]. Однако на практике при внедрении интегрированного управления качеством данных неизбежно возникают трудности, затрагивающие различные аспекты.

На основе вышеописанного подхода авторами разработана верхнеуровневая схема потоков данных между слоями системы на основе хранилища Data Lakehouse [25]. Представленная на рис. 4 схема отображает взаимодействие ключевых компонентов данного хранилища.

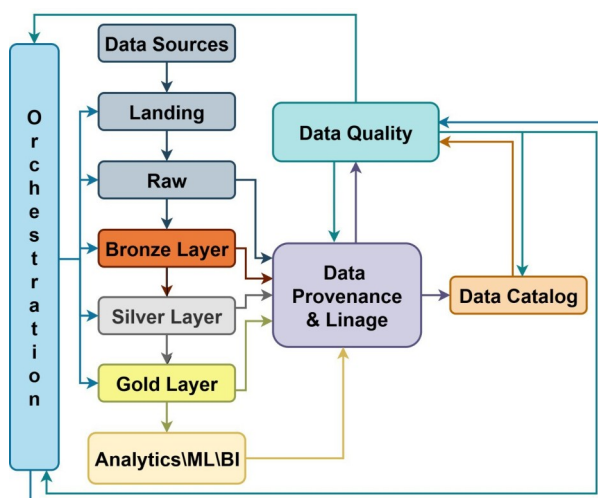


Рис. 4.а

Проведенный в сочетании с практическим опытом авторов анализ позволил выявить ряд критически важных направлений, которые слабо развиты, игнорируются, не соблюдаются или недостаточно контролируются в организациях:

1. Отсутствие понимания взаимосвязи между качеством данных, качеством процессов и информационных систем. Это приводит к недостаточной систематизации и неполноте в определении процессов и систем, оказывающих влияние на качество данных.
2. Недостаточное осознание важности качества данных на уровне руководства, пользователей и ответственных лиц со стороны программы качества данных, из-за чего возникают проблемы во взаимодействии между подразделениями в вопросах качества данных.
3. Недостаток квалифицированных специалистов по качеству данных.
4. Недостаточная проработка стратегии и политики качества данных.
5. Недостаточная интеграция управления качеством данных с другими корпоративными направлениями, системами и хранилищами баз данных.
6. Недооценка экономического эффекта и ROI от улучшения качества данных.
7. Сложности с масштабированием решений по качеству данных ввиду ограниченности систем при росте объема и разнообразия данных.
8. Недостаточная проработка поддержки метаданных, каталогов данных, отслеживания их происхождения и документирования.
9. Отсутствие непрерывного мониторинга и оперативного реагирования на нарушения качества данных.
10. Существенный разрыв между теоретическим потенциалом автоматизации качества данных и существующими решениями. Большинство инструментов сосредоточены на базовой оценке качества без реализации продвинутых механизмов автоматического выявления и применения правил качества данных в хранилищах [30].
11. Отсутствие моделей качества данных, адаптированных для отдельных проектов и бизнес-процессов.
12. Отсутствие единого понимания характеристик качества данных и их мер в рамках организации.
13. В рамках Российской Федерации потенциально недостаточная готовность импортозамещающих систем, поскольку нет достаточной статистической информации об их применении.

5. Заключение

В настоящем исследовании авторами был проведен анализ научных источников, связанных с управлением качеством данных. По результатам анализа были систематизированы стандартные элементы интегрированного управления качеством, описана общая схема и цикл оценки, а также разработана верхнеуровневая схема потоков данных между слоями системы на основе хранилища Data Lakehouse [25]. Предложенные решения рекомендуется учитывать не зависимо от области применения данных.

Можно сделать, что, несмотря на исследовательские инициативы, требуется разработка новых подходов, алгоритмов и методов оптимизации качества больших данных. Новые разработки в части оценки качества позволят выйти за пределы традиционных подходов, сформировавшихся в эпоху преобладания структурированных данных.

Дальнейшие исследования авторов будут сосредоточены на развитии архитектуры, представленной на рис. 4., а также на разработке методов оценки и мониторинга уровня качества данных с учетом характеристик, применяемых для оценки информационных систем, элементов и наборов данных.

Литература

1. Sheng Y. Exploring the mediating and moderating effects of information quality on firm's endeavour on information systems // International Conference on Information Quality (ICIQ 2003). Cambridge, Massachusetts, USA, November 7-9, 2003. P.252–355.
2. Wang R. A product perspective on total data quality management [Электронный ресурс]. URL: <https://dl.acm.org/doi/10.1145/269012.269022> (дата обращения: 21.09.2025).
3. Su Y., Jin Z. A methodology for information quality assessment in the designing and manufacturing processes of mechanical products // International Conference on Information Quality (ICIQ 2004). Cambridge, Massachusetts, USA, November 5-7, 2004. P. 447-465.
4. Kyung-Seok Ryu, Joo-Seok Park, Jae-Hong Park A Data Quality Management Maturity Model // ETRI Journal. 2006. № 28. P. 191–204.
5. Batini C., Scannapieco M. Data Quality: Concepts, Methodologies and Techniques [Электронный ресурс]. URL: <https://dl.acm.org/doi/10.5555/1177291> (дата обращения: 21.09.2025).
6. C. Batini, C. Cappiello, C. Francalanci, A. Maurino Methodologies for data quality assessment and improvement // ACM Computing Surveys. 2009. № 41. P. 1–52.
7. Jeusfeld M., Quix C., Jarke M. Design and analysis of quality information for data warehouses // International Conference on Conceptual Modeling. Berlin, Heidelberg, Germany, November 16-19, 1998. P. 349–362.
8. Chapman A., Richards H., Hawken S. Data and information quality at the Canadian institute for health information [Электронный ресурс]. URL: <http://mitiq.mit.edu/ICIQ/Documents/IQ%20Conference%202006/Papers/Data%20and%20Information%20Quality%20at%20the%20Canadian%20Institute%20for%20Health%20Information.pdf> (дата обращения: 20.09.2025).
9. Eppler MJ., Muenzenmayer P. Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology [Электронный ресурс] URL: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b7e04978992851255d26fd8a00b6673ea9f27f84> (дата обращения: 20.09.2025).
10. Falorsi P., Pallara S., Pavone A., Alessandrini A., Massella E., Scannapieco M. Improving the quality of toponymic data in the italian public administration. In Proceedings of the ICDT Workshop on Data Quality in Cooperative Information Systems (DQCIS) [Электронный ресурс] URL: <https://www.istat.it/> (дата обращения: 20.09.2025).
11. Su Y. and Jin Z. A methodology for information quality assessment in the designing and manufacturing processes of mechanical products // International Conference on Information Quality (ICIQ 2004). Cambridge, Massachusetts, USA, November 5-7, 2004. P. 447–465.
12. Loshin, D. Enterprise Knowledge Management – The Data Quality Approach [Электронный ресурс] URL: <https://dl.acm.org/doi/10.5555/362436> (дата обращения: 20.09.2025).
13. Scannapieco M., Virgillito A., Marchetti M., Mecella M., Baldoni R. The DaQuinCIS architecture: a platform for exchanging and improving data quality in Cooperative Information Systems // Inform. Syst. 2004. № 29. P. 551–582.
14. De Amicis F., Batini C. A methodology for data quality assessment on financial data [Электронный ресурс]. URL: <https://cir.nii.ac.jp/crid/1570291224348968704> (дата обращения: 20.09.2025).
15. The Global Data Management Community [Электронный ресурс]. URL: <https://www.dama.org/cpages/body-of-knowledge> (дата обращения: 20.09.2025).

16. International organization for standardization [Электронный ресурс]. URL: <https://www.iso.org/standard/81745.html> (дата обращения: 20.09.2025).
17. International organization for standardization, International electrotechnical commission [Электронный ресурс]. URL: <https://www.iso.org/standard/78914.html> (дата обращения: 20.09.2025).
18. International organization for standardization, International electrotechnical commission [Электронный ресурс]. URL: <https://www.iso.org/standard/34343.html> (дата обращения: 20.09.2025).
19. International organization for standardization, International electrotechnical commission [Электронный ресурс]. URL: <https://www.iso.org/ru/standard/35736.html> (дата обращения: 20.09.2025).
20. International organization for standardization, International electrotechnical commission [Электронный ресурс]. URL: <https://www.iso.org/standard/81088.html> (дата обращения: 20.09.2025).
21. Центр компетенций НТИ по большим данным МГУ [Электронный ресурс]. URL: <https://bigdata.msu.ru/standards/> (дата обращения: 20.09.2025).
22. Monte Carlo Data [Электронный ресурс]. URL: <https://www.montecarlodata.com/blog-data-quality-testing/> (дата обращения: 20.09.2025).
23. Monte Carlo Data [Электронный ресурс]. URL: <https://www.montecarlodata.com/use-cases/data-quality-monitoring-testing/> (дата обращения: 20.09.2025).
24. Monte Carlo Data [Электронный ресурс]. URL: <https://www.montecarlodata.com/product/data-observability-platform/> (дата обращения: 20.09.2025).
25. Pure Storage Blog [Электронный ресурс]. URL: <https://blog.purestorage.com/purely-educational/data-fabric-vs-data-lake-vs-data-warehouse/> (дата обращения: 20.09.2025).
26. Ofner M., Otto B., Osterle H. Integrating a data quality perspective into business process management // Business Process Management Journal. 2012. № 18. P. 1036-1067.
27. Al-Sai Z., Gandomi A., Al-Sai Z., Al-Nuaimi E., Al-Jaroodi J., Mohamed N., Al-Neyadi H., Al-Bayati A., Al-Kahtani M. Big Data Maturity Assessment Models: A Systematic Literature Review // Big Data Cognition and Computation. 2023. № 7. P. 1-27.
28. The DGI Data Governance [Электронный ресурс]. URL: <https://datagovernance.com/> (accessed: 20.09.2025).
29. IBM Data Governance Framework [Электронный ресурс]. URL: <https://www.ibm.com/products/cloud-pak-for-data/governance> (дата обращения: 20.09.2025).
30. Heidi C., Nikiforova A. Towards augmented data quality management: Automation of Data Quality Rule Definition in Data Warehouses [Электронный ресурс]. URL: <https://arxiv.org/abs/2406.10940> (дата обращения: 20.09.2025).
31. Hazen B., Boone C., Ezell J., Jones-Farmer L. Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications // International Journal of Production Economics. 2014. № 154. P. 72-80.
32. Leghemo I., Osinachi D., Chinekwu S., Chima A. Continuous Data Quality Improvement in Enterprise Data Governance: A Model for Best Practices and Implementation // Engineering Research and Reports. 2025. № 27. P. 29-25.
33. Taleb I., Serhani M., Bouhaddioui C., Dssouli R. 2021. Big data quality framework: a holistic approach to continuous quality management // Journal of Big Data. 2021. № 8. P. 1–41.

34. Bello H., Ige A., Ameyaw M. Deep learning in high-frequency trading: conceptual challenges and solutions for real-time fraud detection // World Journal of Advanced Engineering Technology and Sciences. 2024. № 12, P. 35-46.
35. Cappiello C., Cerletti C., Fratto C., and Pernici B. Validating data quality actions in scoring processes. Journal of Data and Information Quality (JDIQ). 2018. № 9. P. 1–27.

Захарова Оксана Игоревна

к.т.н., доцент кафедры информационных систем и технологий, Поволжский государственный университет телекоммуникаций и информатики (ПГУТИ, 443010, Самара, ул. Льва Толстого, д. 23), тел. +7 906 343 2521, e-mail: o.zaharova@psuti.ru, ORCID ID: 0000-0003-3371-4344.

Коробейников Владислав Сергеевич

аспирант, Поволжский государственный университет телекоммуникаций и информатики (ПГУТИ, 443010, Самара, ул. Льва Толстого, д. 23), тел. +7 987 952 1530, e-mail: vlad.k.k78@gmail.com, ORCID ID: 0009-0008-7937-5476.

Авторы прочитали и одобрили окончательный вариант рукописи.

Авторы заявляют об отсутствии конфликта интересов.

Вклад соавторов: Каждый автор внес равную долю участия как во все этапы проводимого теоретического исследования, так и при написании разделов данной статьи.

Integrated data quality management within the organization

Oksana I. Zaharova, Vladislav S. Korobeinikov

¹Povolzhskiy State University of Telecommunications and Information Science (PSUTI)

Abstract: The relevance of integrated data quality management tasks is increasing in the context of growing volume, variety, and criticality of data used. Despite this, organizations still have significant gaps in understanding the interconnections between data quality, process quality, and information systems. The purpose of this study is a systematic analysis of existing methodologies and concepts for data quality management, as well as the identification of key challenges in their practical implementation. The paper presents a review of scientific literature, outlines standard elements and a framework for an integrated approach to data quality. A high-level data flow scheme based on the Data Lakehouse architecture has been developed, reflecting the interaction of system components. The necessity of developing new methods and algorithms for optimizing big data quality, which go beyond traditional paradigms focused on structured data, is substantiated. Key problems often ignored in practice are identified and systematized, and criteria for the successful implementation of an integrated data quality management approach are formed.

Keywords: Data Quality, Data Quality Management, Integrated Data Quality Management, Data Quality Methodology, Data Quality Monitoring, Big Data, Information Quality.

For citation: Zaharova O. I., Korobeinikov V. S. Integrated data quality management within the organization [Paper Preparation Manual for Vestnik SibGUTI]. Vestnik SibGUTI, 2025, vol. 19, no. 4, pp. 28-47. <https://doi.org/10.55648/1998-6920-2025-19-4-28-47>.



Content is available under the license
Creative Commons Attribution 4.0
License

© Zaharova O. I., Korobeinikov V. S., 2025

The article was submitted: 24.09.2025;
revised version: 30.09.2025;
accepted for publication 21.10.2025.

References

1. Sheng Y. Exploring the mediating and moderating effects of information quality on firm's endeavour on information systems. International Conference on Information Quality. Cambridge, November 7-9, 2003. pp. 252–355.
2. Wang R. A product perspective on total data quality management, available at <https://dl.acm.org/doi/10.1145/269012.269022> (accessed: 02.05.2025).
3. Su Y., Jin Z. A methodology for information quality assessment in the designing and manufacturing processes of mechanical products // International Conference on Information Quality. Cambridge, November 5-7, 2004. pp. 447-465.
4. Kyung-Seok Ryu, Joo-Seok Park, Jae-Hong Park A Data Quality Management Maturity Model. ETRI Journal. 2006. no. 28. pp. 191–204.
5. Batini C., Scannapieco M. Data Quality: Concepts, Methodologies and Techniques, available at <https://dl.acm.org/doi/10.5555/1177291> (accessed: 02.05.2025).
6. C. Batini, C. Cappiello, C. Francalanci, A. Maurino Methodologies for data quality assessment and improvement // ACM Computing Surveys. 2009. no. 41, pp. 1–52.
7. Jeusfeld M., Quix C., Jarke M. Design and analysis of quality information for data warehouses. International Conference on Conceptual Modeling. Heidelberg, Germany, November 16-19, 1998. pp. 349–362.
8. Chapman A., Richards H., Hawken S. Data and information quality at the Canadian institute for health information, available at <http://mitiq.mit.edu/ICIQ/Documents/IQ%20Conference%202006/Papers/Data%20and%20Information%20Quality%20at%20the%20Canadian%20Institute%20for%20Health%20Information.pdf> (accessed: 02.05.2025).
9. Eppler MJ., Muenzenmayer P. Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology, available at <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b7e04978992851255d26fd8a00b6673ea9f27f84> (accessed: 02.05.2025).
10. Falorsi P., Pallara S., Pavone A., Alessandrini A., Massella E., Scannapieco M. Improving the quality of toponymic data in the Italian public administration. In Proceedings of the ICDT Workshop on Data Quality in Cooperative Information Systems (DQCIS), available at <https://www.istat.it/> (accessed: 02.05.2025).
11. Su Y. and Jin Z. A methodology for information quality assessment in the designing and manufacturing processes of mechanical products. International Conference on Information Quality. Cambridge, November 5-7, 2004, pp. 447–465.
12. Loshin, D. Enterprise Knowledge Management – The Data Quality Approach, available at <https://dl.acm.org/doi/10.5555/362436> (accessed: 02.05.2025).
13. Scannapieco M., Virgillito A., Marchetti M., Mecella M., Baldoni R. The DaQuinCIS architecture: a platform for exchanging and improving data quality in Cooperative Information Systems. Inform. Syst. 2004. no. 29, pp. 551–582.
14. De Amicis F., Batini C. A methodology for data quality assessment on financial data, available at <https://cir.nii.ac.jp/crid/1570291224348968704> (accessed: 02.05.2025).
15. The Global Data Management Community, available at <https://www.dama.org/cpages/body-of-knowledge> (accessed: 02.05.2025).
16. International organization for standardization, available at <https://www.iso.org/standard/81745.html> (accessed: 02.05.2025).
17. International organization for standardization, International electrotechnical commission, available at <https://www.iso.org/standard/78914.html> (accessed: 02.05.2025).

18. International organization for standardization, International electrotechnical commission, available at <https://www.iso.org/standard/34343.html> (accessed: 02.05.2025).
19. International organization for standardization, International electrotechnical commission, available at <https://www.iso.org/ru/standard/35736.html> accessed: 02.05.2025).
20. International organization for standardization, International electrotechnical commission, available at <https://www.iso.org/standard/81088.html> (accessed: 02.05.2025).
21. NTI Competence Center according to large data from Moscow State University, available at <https://bigdata.msu.ru/standards/> (accessed: 02.05.2025).
22. Monte Carlo Data, available at <https://www.montecarlodata.com/blog-data-quality-testing/> (accessed: 02.05.2025).
23. Monte Carlo Data, available at <https://www.montecarlodata.com/use-cases/data-quality-monitoring-testing/> (accessed: 02.05.2025).
24. Monte Carlo Data, available at <https://www.montecarlodata.com/product/data-observability-platform/> (accessed: 02.05.2025).
25. Pure Storage Blog, available at <https://blog.purestorage.com/purely-educational/data-fabric-vs-data-lake-vs-data-warehouse/> (accessed: 02.05.2025).
26. Ofner M., Otto B., Osterle H. Integrating a data quality perspective into business process management. *Business Process Management Journal*. 2012. no 18. pp. 1036-1067.
27. Al-Sai Z., Gandomi A., Al-Sai Z., Al-Nuaimi E., Al-Jaroodi J., Mohamed N., Al-Neyadi H., Al-Bayati A., Al-Kahtani M. Big Data Maturity Assessment Models: A Systematic Literature Review. *Big Data Cognition and Computation*. 2023. no 7. pp. 1-27.
28. The DGI Data Governance, available at <https://datagovernance.com/> (accessed: 02.05.2025).
29. IBM Data Governance Framework, available at <https://www.ibm.com/products/cloud-pak-for-data/governance> (accessed: 02.05.2025).
30. Heidi C., Nikiforova A. Towards augmented data quality management: Automation of Data Quality Rule Definition in Data Warehouses, available at URL: <https://arxiv.org/abs/2406.10940> (accessed: 02.05.2025).
31. Hazen B., Boone C., Ezell J., Jones-Farmer L. Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*. 2014. no 154. pp. 72-80.
32. Leghemo I., Osinachi D., Chinekwu S., Chima A. Continuous Data Quality Improvement in Enterprise Data Governance: A Model for Best Practices and Implementation. *Engineering Research and Reports*. 2025. no 27. pp. 29-25.
33. Taleb I., Serhani M., Bouhaddioui C., Dssouli R. 2021. Big data quality framework: a holistic approach to continuous quality management. *Journal of Big Data*. 2021. no. 8, pp. 1–41.
34. Bello H., Ige A., Ameyaw M. Deep learning in high-frequency trading: conceptual challenges and solutions for real-time fraud detection. *World Journal of Advanced Engineering Technology and Sciences*. 2024. no. 12, pp. 35-46.
35. Cappiello C., Cerletti C., Fratto C., and Pernici B. Validating data quality actions in scoring processes. *Journal of Data and Information Quality (JDIQ)*. 2018. № 9. P. 1–27.

Oksana I. Zaharova

PhD. (Engineering), Associate Professor; Department of Information Systems and Technologies, Povolzhskiy State University of Telecommunications and Information Science (PSUTI, Russia, 443010, Samara, Lev Tolstoy St. 23), phone: +7 987 952 1530, e-mail: o.zaharova@psuti.ru, ORCID ID: 0000-0003-3371-4344.

Vladislav S. Korobeinikov

PhD student, Povolzhskiy State University of Telecommunications and Information Science (PSUTI, Russia, 443010, Samara, Lev Tolstoy St. 23), phone: +7 987 952 1530, e-mail: vlad.k.k78@gmail.com, ORCID ID: 0009-0008-7937-5476.